Journal of Biogeography (*J. Biogeogr.*) (2007) **34**, 736–752

ORIGINAL ARTICLE

Plastid and nuclear DNA polymorphism reveals historical processes of isolation and reticulation in the olive tree complex (*Olea europaea*)

G. Besnard[1]*, R. Rubio de Casas[2,3] and P. Vargas[3]

[1]*Department of Ecology and Evolution, Biophore, University of Lausanne (UNIL), 1015 Lausanne, Switzerland,* [2]*Departamento de Biología Vegetal 1, Universidad Complutense de Madrid (UCM), José Antonio Novais 2, 28040 Madrid, Spain and* [3]*Royal Botanic Garden of Madrid, Consejo Superior de Investigaciones Científicas (CSIC), Plaza de Murillo 2, 28014 Madrid, Spain*

*Correspondence: G. Besnard, Department of Ecology and Evolution, Biophore, University of Lausanne, 1015 Lausanne, Switzerland.
E-mail: gbesnard@unil.ch.

ABSTRACT

Aim The olive tree is considered one of the best indicators of the Mediterranean climate. The species' distribution is associated with geographical and bioclimatic factors, as well as being influenced by a long period of cultivation. Despite concerted efforts of different research groups, the origin of the Mediterranean olive tree still remains elusive. In the present study, relationships between taxa and populations covering the entire range of *Olea europaea* were investigated using both maternal (plastid genome) and biparental (nuclear genome) markers to disclose evolutionary patterns in the olive complex. Phylogenetic and phylogeographical results of the two-genome analyses were interpreted in a biogeographical context.

Location Mediterranean, temperate and subtropical floristic regions of the Old World.

Methods Phylogeographical reconstructions of plastid DNA polymorphism were performed using microsatellites, restriction sites and indels on a wide sample of 185 representative trees across the Old World, including 28 herbarium specimens from remote areas. Additionally, the potential utility of one ITS-1 pseudogene for phylogenetic analyses was explored using Bayesian and maximum parsimony approaches on a subsample of 38 olive trees.

Results Forty plastid haplotypes were recognized and split into two lineages and seven sublineages. The analysis of ITS-1 sequences also allowed the identification of seven well differentiated groups. Distribution of plastid and ribosomal DNA lineages was congruent, but particular cases of phylogenetic incongruence were disclosed (particularly in the Sahara and Madeira). Lastly, two divergent ITS-1 copies were isolated from the same sample of four individuals of different subspecies.

Main conclusions Phylogenetic congruence of both ITS-1 and plastid lineages suggested an evolutionary scenario of predominant isolation during the Plio-Pleistocene in Macaronesia, the Mediterranean, southern Africa, eastern Africa and Asia. The Saharan desert appeared to have played an important role of vicariant barrier between southern and northern African populations in early times. Incongruence of some plastid and nuclear results, as well as intermingled ITS-1 copies of different lineages in single individuals, was interpreted as a result of recurrent reticulation events in the olive complex. We identified an ancient hybrid zone from the Sahara to north-eastern African mountains, where divergent plastid and nuclear lineages still co-exist. Results of this paper, and previous studies, suggest that the cultivated olive originated from a pre-Quaternary Mediterranean ancestor, with no evidence for a recent hybrid origin. In contrast, a continuous process of olive domestication through local hybridization events of cultivated trees with natural populations may have brought about a remarkably high genomic diversity among cultivated trees across the Mediterranean.

doi:10.1111/j.1365-2699.2006.01653.x

Keywords
Evolutionary patterns, hybridization, internal transcribed spacer, Mediterranean region, olive, phylogeography, plastid DNA, pseudogenes.

INTRODUCTION

The biological origin of domesticated plants is increasingly being clarified by means of phylogeographical and phylogenetic reconstructions (Armelagos & Harper, 2005; Zeder *et al.*, 2006). Despite the concerted effort of different laboratories (Angiolillo *et al.*, 1999; Besnard *et al.*, 2001a, 2002; Contento *et al.*, 2002; Owen *et al.*, 2005; Breton *et al.*, 2006; Rubio de Casas *et al.*, 2006), the precise origin of the domesticated Mediterranean olive tree (*Olea europaea* L. subsp. *europaea*) has remained elusive. Relatives of the Mediterranean olive tree are circumscribed in five subspecies (Green, 2002): *laperrinei*, distributed in Saharan massifs (Hoggar, Aïr, Jebel Marra); *cuspidata*, from South Africa to southern Egypt and from Arabia to northern India and south-west China; *guanchica* in the Canary Islands; *maroccana* in south-western Morocco; and *cerasiformis* in Madeira. Four historical hypotheses have been formulated to account for the origin of the Mediterranean olive tree on the basis of the geographical distribution of morphotypes and chemotypes: (1) migration, establishment and evolution from populations of subsp. *cuspidata* occurring in Tropical Africa or southern Asia; (2) isolation of populations of subsp. *laperrinei* in the Mediterranean after Quaternary range contractions; (3) hybridization between two taxa (related to subsp. *europaea* and *cuspidata*) co-occurring in eastern Mediterranean mountains; and (4) divergence from an *O. europaea* lineage in the Mediterranean (Oliver, 1868; Newberry, 1937; Chevalier, 1948; Turrill, 1951; Elant, 1976; Green & Wickens, 1989; Zohary & Hopf, 2000). Palaeobotanical evidence also suggests that the past distribution of populations of subspecies *laperrinei* (Batt. & Trab.) Ciferri and *cuspidata* was adjacent to the Mediterranean basin (Quézel, 1978). The question remains whether these two taxa contributed to the evolution of the Mediterranean olive in the Pliocene and Pleistocene, much earlier than the present-day pattern of extensive hybridization between wild and cultivated olives throughout the Mediterranean (Zohary & Hopf, 2000; Vargas & Kadereit, 2001; Besnard *et al.*, 2002).

Common maternal inheritance of organelles makes cytoplasmic DNA polymorphism an ideal tool for phylogeographical study and assessment of phylogenetic relationships between uniparental lineages disseminated by seeds (Schaal *et al.*, 1998; McKinnon *et al.*, 2004; Petit *et al.*, 2005). Cytoplasmic DNA polymorphism was used to evaluate the origin of wild Mediterranean and cultivated olives (Besnard *et al.*, 2000, 2002; Vargas & Kadereit, 2001; Baldoni *et al.*, 2002). To date, 15 plastid DNA haplotypes from five main clades corresponding to specific geographical areas (clade A, eastern and southern Africa; clade C, southern Asia; clade E1, Saharan Africa and the Mediterranean; clade E2, western Mediterranean; clade M, southern Morocco and the Canary Islands) have been identified in the olive complex (Besnard *et al.*, 2002). This clear-cut phylogeographical structure is congruent with palaeobotanical evidence for evolution of 'modern' lineages during the Plio-Pleistocene (Terral *et al.*, 2004b). Populations of *O. europaea* from eastern Africa were underrepresented in most studies because of sampling difficulties, even though this geographical area has been hypothesized as an important centre of diversity for *O. europaea* (Oliver, 1868; Besnard *et al.*, 2002). Incongruence between fingerprinting results used to analyse relationships between *O. europaea* taxa left open the question whether hybridization had taken place in eastern Africa (Angiolillo *et al.*, 1999; Hess *et al.*, 2000; Besnard *et al.*, 2001b). Results based solely on either maternal or biparental markers are, however, insufficient to infer reticulate evolution in any species complex (Rieseberg *et al.*, 1996).

The analysis of both cytoplasmic and nuclear markers, particularly the use of maternally and biparentally inherited markers, can help to detect hybridization between differentiated populations, and to reconstruct ancestral lineage-sorting events during species formation (Comes & Abbott, 2001; Cronn *et al.*, 2003; Doyle *et al.*, 2003; Chat *et al.*, 2004; Cronn & Wendel, 2004). The internal transcribed spacers (ITS) of the nuclear ribosomal DNA (nrDNA) are biparentally-inherited markers, and have been used extensively in the past decade in the analysis of angiosperms, because they display informative polymorphism and can be readily amplified and sequenced even from poorly preserved material (Baldwin *et al.*, 1995; Álvarez & Wendel, 2003). Due to a high number of ribosomal DNA copies in a genome, ITS sequences can also be recovered from archaeological material (Elbaum *et al.*, 2006). Ribosomal genes are subjected to concerted evolution, potentially leading to uniformity of ITS sequences at the individual, population and species levels (Liao, 2003; Rauscher *et al.*, 2004; Kovarik *et al.*, 2005). When failure of concerted evolution occurs, the ITS region allows detection of hybrids containing multiple ITS copies (Ainouche & Bayer, 1997; Vargas *et al.*, 1999; Manen, 2004; Ritz *et al.*, 2005). Although ITS sequences were informative in the *O. europaea* complex, technical difficulties (multiple DNA amplicons) were encountered when sequencing ITS and external transcribed spacer (ETS) regions (Hess *et al.*, 2000). Pseudogenes have been reported in numerous species, and can lead to additional methodological difficulties in phylogeographical analysis based on divergent copies (Buckler *et al.*, 1997; Bailey *et al.*, 2003). Nevertheless, pseudogenes of

the same type, similarly to active genes, are useful for phylogenetic inferences, particularly when the functional counterparts are not variable (Razafimandimbison *et al.*, 2004). In any case, a detailed characterization of nrDNA orthology and paralogy is essential (Bailey *et al.*, 2003).

Restricted sampling in previous studies (Vargas & Kadereit, 2001; Baldoni *et al.*, 2002; Besnard *et al.*, 2002), coupled with reconstructions based on single genomic markers, contributed to an incomplete understanding of the evolution of the *O. europaea* complex and the origin of domesticated olive trees. In this sense, herbarium samples from remote areas become crucial in order to screen regions previously not prospected. In the present study, molecular analyses of a large number of plastid haplotypes and ribosomal DNA sequences were performed to infer the phylogenetic relationships between the six subspecies of the *O. europaea* complex, and the phylogeographical structure of populations. Lastly, analysis of plastid and nuclear ribosomal DNA data allowed reconstruction of evolutionary patterns (isolation vs. hybridization) involved in differentiation of the olive complex.

MATERIALS AND METHODS

Sampling strategy and characterization of plastid DNA polymorphism

A total of 157 wild and cultivated individuals were sampled in the field (see Supplementary Material). DNA from fresh leaves desiccated in silica gel was extracted according to Besnard *et al.* (2000). In addition, 28 herbarium samples, collected in 20 sites from 15 countries, were characterized to take into account individuals from remote areas never analysed so far (particularly from East Africa; see Supplementary Material). DNA of herbarium material was extracted from 20 mg leaf tissue using the DNeasy Plant Mini Kit (Qiagen, GmbH, Germany). To verify reproducibility of the plastid DNA polymorphism characterization, two independent DNA extractions were performed for four herbarium samples (A. Faure s.n., Newberry 2, Fahmy & Hassib s.n. and Maas Geesteranus 5998). In total, 16 cultivated and ornamental trees and 169 wild individuals from 62 populations were analysed.

Sequence polymorphism of the plastid DNA (in *trn*T-L, *trn*S-G and *mat*K spacers) was initially screened in 20 individuals from the six subspecies (Besnard *et al.*, 2003), but sequence variation was particularly low (minimum 99.7% similarity between two haplotypes). Once the most informative plastid regions were identified, we analysed polymorphic microsatellites, indels and cleaved amplified polymorphism (CAP) loci because they: (1) can be used on forensic material; (2) allow screening of polymorphism on independent spacers; and (3) can be extended to large numbers of samples. A crucial point in molecular characterization was the PCR amplification of short DNA fragments (< 120 bp) to screen poorly preserved material (herbarium specimens). The samples were screened for eight polymorphic plastid DNA multi-state microsatellite motifs (Weising & Gardner, 1999; Besnard *et al.*, 2003; Baali-Cherif & Besnard, 2005), two deletions of 8 and 10 bp (Besnard *et al.*, 2003), and 11 CAP markers (Besnard *et al.*, 2003; Baali-Cherif & Besnard, 2005; Table 1). These markers were combined to obtain a multi-locus plastid DNA haplotype for each individual.

Characterization of the ITS-1 spacer

To incorporate biparental data, a subsample of 38 individual DNAs, including six herbarium specimens of different subspecies, a wide range of geographical distribution, and plastid lineages, was sequenced for ITS-1 (Table 2; Fig. 4). Technical difficulties (due to simultaneous PCR amplification of divergent ITS copies) had already been observed in the olive tree complex when generating ITS sequences with different PCR conditions and several angiosperm pairs of primers (ITS1-For and ITS4-Rev, Brown *et al.*, 1996 and A. Moukhli, unpublished data; 17SE and 26SE, Hess *et al.*, 2000; ITS-A and ITS-C, A. Moukhli, unpublished data). Hess *et al.* (2000) successfully isolated ITS-1 sequences from five subspecies. These ITS-1 accessions (AF251360–AF251364) were used to design one forward internal ITS-1 primer (ITS-1f: 5′-AAA-AGGTAGACCCAYGAACTCG-3′) to prevent amplification of double bands and to amplify exclusively the same ribosomal DNA-sequence copy retrieved by Hess *et al.* (2000). We also designed one reverse primer (5.8Sr: 5′-TCGC-ATTTTGCTGCGTTCTTC-3′) located at the 5′ end of the 5.8S nrDNA. The PCR reaction mixture contained 1–10 ng DNA template, 1x reaction buffer, 2 mM $MgCl_2$, 0.2 mM dNTPs, 0.2 μmol of each primer, and 0.75 U DNA polymerase (AccuPrime Taq, Invitrogen, Carlsbad, CA, USA) in a total volume of 25 μL. After 4 min at 94°C, the PCR program in a thermocycler (T1, Biometra, Göttingen, Germany) was: 36 cycles of 45 s at 94°C, 45 s at 53°C, and 45 s at 72°C. The last cycle was followed by a 10-min extension at 72°C. Direct sequencing was performed using the Big Dye 3.1 Terminator cycle-sequencing kit (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's instructions and with an ABI Prism 3100 genetic analyser (Applied Biosystems). Despite amplifying a single band in all cases, double-peak chromatograms were observed in 17 individuals. For these samples, ITS-1 haplotypes were isolated using the InsT/Aclone PCR product cloning kit (Mbi Fermentas Inc., ON, Canada). We screened positive bacterial colonies characterized by an insert of *c.* 260–270 bp. For each clone, the insert was directly PCR-amplified using M13 primers (using the PCR conditions mentioned above) and then separated on a 3% agarose gel to estimate fragment size. Ten positive transformed bacterial colonies were isolated per individual. Each clone was digested with *Taq*I (Mbi Fermentas). The degree of polymorphism for *Taq*I was particularly high in ITS-1 sequences from Hess *et al.* (2000), and allowed an unambiguous distinction of all identified lineages (see below). Restriction fragments were separated on a 3% agarose gel, and different clones (revealed by different profiles) were sequenced directly as previously described. This methodology (direct sequencing and cloning followed by

Table 1 Multi-locus profiles of the 40 plastid DNA haplotypes obtained from 185 trees of the *Olea europaea* complex

		Plastid DNA variation																				
PtDNA haplotype	Haplotype code*	*ccmp5*†	*ccmp7*†	*trn*T-L-polyT	*trn*T-L-polyA	*psb*K-*trn*S-polyT-A	*trn*G-polyT	*trn*G-R-polyT	*trn*S-G-indel-1‡	*trn*S-G-indel-2	*trn*T-L-*Taq*I	*trn*T-L-*Acs*I	*mat*K-*Rsa*I	*mat*K-*Taq*I	*mat*K2-*Taq*I	*mat*K2-*Mse*I	*mat*K3-*Taq*I	*trn*S-G-*Mse*I	*trn*S-G-*Nde*I	*TrnG*-*Pst*I	*psb*K-*trn*S-*Mse*I
1	CE1	105	120	81	106	109	87	60	105	60	+	−	+	+	−	−	+	+	+	−	−
2	CE2	106	120	81	106	109	87	60	105	60	+	−	+	+	−	−	+	+	+	−	−
3	CL1	105	120	81	106	109	87	60	105	60	+	−	+	+	−	−	+	+	−	−	−
4	CL2	105	120	81	106	108	87	60	105	60	+	−	+	+	−	−	+	+	−	−	−
5	CL3	105	120	82	106	109	87	60	105	60	+	−	+	+	−	−	+	+	−	−	−
6	CL4	104	120	81	106	109	87	60	105	60	+	−	+	+	−	−	+	+	−	−	−
7		107	120	81	106	109	87	60	105	60	+	−	+	+	−	−	+	+	−	−	−
8		106	120	82	106	110	87	60	105	60	+	−	+	+	−	−	+	+	−	−	−
9	CCK	106	120	81	106	109	87	60	114	60	+	−	+	+	−	−	+	+	−	−	−
10	CCK2	105	120	81	106	109	87	60	114	60	+	−	+	+	−	−	+	+	−	−	−
11	COM1	104	121	81	106	108	87	60	117	52	+	−	+	+	−	−	+	−	−	−	−
12	COM2	105	121	81	106	108	87	60	117	52	+	−	+	+	−	−	+	−	−	−	−
13		104	121	81	106	108	87	60	117	53	+	−	+	+	−	−	+	−	−	−	−
14	CCE1	106	120	80	106	109	87	60	115	60	+	−	+	+	−	−	+	+	−	−	+
15	CCE2	107	120	80	106	109	87	60	115	60	+	−	+	+	−	−	+	+	−	−	+
16	CCE3	108	120	80	106	109	87	60	115	60	+	−	+	+	−	−	+	+	−	−	+
17		107	120	80	106	110	87	60	115	60	+	−	+	+	−	−	+	+	−	−	+
18		107	120	80	106	110	88	60	115	60	+	−	+	+	−	−	+	+	−	−	+
19		107	120	80	106	109	87	60	115	60	+	−	+	+	−	−	−	+	−	−	+
20		108	120	80	106	109	87	60	115	60	+	−	+	+	−	−	−	+	−	−	+
21		106	120	80	106	110	87	60	115	60	+	−	+	+	−	−	+	+	−	−	+
22		105	120	80	106	111	87	60	115	60	+	−	+	+	−	−	+	+	−	−	+
23		105	120	80	106	110	87	60	115	60	+	−	+	+	−	−	+	+	−	−	+
24	CC1	105	119	79	106	108	86	60	114	60	−	−	−	−	+	−	+	+	−	+	−
25		105	119	80	106	108	86	60	114	60	−	−	−	−	+	−	+	+	−	+	−
26		105	119	79	106	109	86	60	114	60	−	−	−	−	+	−	+	+	−	+	−
27	CC2	103	119	80	107	109	86	60	114	60	−	−	−	−	+	−	+	+	−	+	−
28		103	119	81	107	109	86	60	114	60	−	−	−	−	+	−	+	+	−	+	−
29		103	119	80	107	108	86	60	114	60	−	−	−	−	+	−	+	+	−	+	−
30		102	119	80	106	109	86	60	114	60	−	−	−	−	+	−	+	+	−	+	−
31		103	120	81	106	109	86	60	114	60	−	−	−	−	+	−	+	+	−	+	−
32		103	119	81	106	109	86	60	114	60	−	−	−	−	+	−	+	+	−	+	−
33		104	120	79	106	109	86	60	114	60	−	−	−	−	+	−	+	+	−	+	−
34		104	120	79	106	108	86	60	114	60	−	−	−	−	+	−	+	+	−	+	−
35	CA1	105	120	79	106	108	86	61	114	60	+	−	−	−	+	+	+	+	−	+	−
36	CA2/CA3	106	120	79	106	108	86	61	114	60	+	−	−	−	+	+	+	+	−	+	−
37	CA4	105	120	79	106	108	86	60	114	60	+	−	−	−	+	+	+	+	−	+	−

Table 1 *continued*

PtDNA haplotype	Haplotype code*	Plastid DNA variation																	
		ccmp5†	*ccmp7*†	*trn*T-L-polyT	*trn*T-L-polyA	*psb*K-*trn*S-polyT-A	*trn*G-polyT	*trn*G-R-polyT	*trn*S-G-indel-1‡	*trn*S-G-indel-2	*trn*T-L-*Taq*I	*trn*T-L-*Acs*I	*mat*K-*Rsa*I	*mat*K-*Taq*I	*mat*K2-*Taq*I	*mat*K2-*Mse*I	*mat*K3-*Taq*I	*trn*S-G-*Mse*I	*trn*S-G-*Nde*I	*Trn*G-*Pst*I	*psb*K-*trn*S-*Mse*I
38	CA5	105	120	79	106	109	86	61	114	60	+	–	–	–	+	+	+	+	–	+	–
39		105	120	79	106	108	86	61	114	60	+	+	–	–	+	+	+	+	–	+	–
40		105	120	79	106	108	86	61	115	60	+	–	–	–	+	+	+	+	–	+	–

Fragment sizes (bp) are given for the first nine markers (microsatellites and indels). Restriction site presence (+) or absence (–) is indicated for cleaved amplified polymorphisms (CAPs).
*Haplotype code as defined by Besnard *et al.* (2002); Baali-Cherif & Besnard (2005).
†Fragment size was corrected in comparison with the previous estimation (Besnard *et al.*, 2002).
‡Two distinct characters were considered separately in the phylogenetic analysis: a 10-bp indel and the length variation of a poly-T motif (Besnard *et al.*, 2003).

PCR–RFLP screening) allowed efficient detection of divergent ITS-1 copies within a single individual.

Data analysis

To reconstruct a plastid DNA genealogy of the *O. europaea* complex, a reduced median network was built based on the CAP (restriction site presence or absence, indels) and length multi-state of microsatellites (Bandelt *et al.*, 1999; Posada & Crandall, 2001). This maximum-parsimony analysis was performed using the NETWORK software (http://www.fluxus-engineering.com/sharenet.htm). Recurrent mutations are more frequent on microsatellite motifs than those observed on substitutions (Provan *et al.*, 1999). In fact, size homoplasy in *O. europaea* has been inferred previously at two microsatellite loci (ccmp5 and *psb*K-*trn*S-polyT/A; Besnard *et al.*, 2002, 2003). To downweight these microsatellite markers, we arbitrarily assigned a weight of two for character changes in both CAPs and non-repeat indels, but no correction was made for haplotype frequencies (Bandelt *et al.*, 1999).

Fungal endophytes or other contaminants of the nrDNA ITS-1 sequences were not identified by BLAST research. All available ITS-1 olive sequences deposited in GenBank (Hess *et al.*, 2000; P. Talhinhas and co-workers, unpublished data; P.V., unpublished data) and those generated in the present study were aligned using CLUSTALW (Thompson *et al.*, 1994). Manual alignments were necessary when including outgroup sequences of other Oleaceae (*Fraxinus*, *Jasminum*, *Ligustrum*, *Osmanthus*, *Syringa*). To help identification of pseudogenes, the minimum-energy secondary structure of ITS-1 of each sequence was estimated with MFOLD (http://mfold.burnet.edu.au; Zuker, 1989) using the default temperature of 37°C.

We used PAUP 4.0b8b (Swofford, 2001) to compute pairwise sequence divergence, frequency of G + C, and maximum parsimony (MP) analysis. Parsimony-based analyses were conducted using unweighted Fitch parsimony (as implemented in PAUP 4.0b10; Swofford, 2001), using a heuristic searches strategy of 1000 random addition replicates followed by tree bisection-reconnection (TBR) branch-swapping and with the options MULPARS and STEEPEST DESCENT in effect. Relative support for clades identified by parsimony was assessed by 1000 bootstrap (bs) resamplings using the heuristic search strategy indicated above.

To determine the model of sequence evolution that best fitted the sequence data, the hierarchical likelihood ratio test was computed with MODELTEST 3.06 (Posada & Crandall, 1998). Among the 56 models implemented in the software, the one chosen for the ITS matrix was the HKY + G model. Bayesian inference (BI) was performed using the same model of substitutions with MRBAYES ver. 3.1.1 (Ronquist & Huelsenbeck, 2003). Bayesian inference was performed with the following settings: the Markov Chain Monte Carlo (MCMC) algorithm was run for 5,000,000 generations, with four concurrent chains (one heated), saving the current tree

Table 2 List of *Olea europaea* individuals analysed using both ITS-1 sequences and plastid DNA haplotypes

Individual	Accession no.	ITS-1 lineage	Plastid lineage†
Subsp. *europaea*			
Adana, Turkey	AJ877270‡, AJ877271‡	III	E1 (1)
Ali no. 10, Sicily, Italy	AJ877282§	II	E2 (12)
Birkhadem no. 7, Algiers, Algeria	AJ877279§	II	E2 (12)
Cadiz, Spain	AM162274§	II	E2 (11)
Cyrenaique no. 8, Libya	AJ877274§	III	E1 (1)
Harem no. 1, Syria	AJ877269§	III	E1 (1)
Mt Belloua no. 5, Kabylie, Algeria	AJ877280§	II	E3 (9)
Ostricone no. 5, Corsica, France	AJ877281§	II	E2 (11)
Tamanar no. 1, Morocco	AJ877278§	II	E2 (11)
Tiznit no. 1, Anti Atlas, Morocco	AJ877276‡, AJ877277‡	II, III	E2 (11)
Torvizcon no. 5, Spain	AJ877273§	III	E1 (1)
Urla no. 2, Turkey	AJ877272§	III	E1 (1)
cv. Olivière, France	AJ877275§	III	E3 (9)
cv. Taksrit, Algeria	AJ877283§	II	E1 (1)
Subsp. *guanchica* P. Vargas *et al.*			
La Gomera no. 1, Canary Islands, Spain	AJ877285‡, AJ877286‡	V	M (17)
La Palma no. 2, Canary Islands, Spain	AJ877284§	V	M (16)
Subsp. *maroccana* (Greut. & Burd.) P. Vargas *et al.*			
Mentaga no. 2, High Atlas, Morocco	AJ877287‡, AJ877288‡	V, VII	M (19)
Immouzzer no. 1, High Atlas, Morocco	AJ877289§	V	M (20)
Subsp. *cerasiformis* G. Kunkel & Sunding			
Arco de Calheta, Madeira, Portugal	AJ877291‡, AJ877292‡	II	M (21)
São Gonçalo, Madeira, Portugal*	AJ877290§	II	M (22)
Subsp. *laperrinei* (Batt. & Trab.) Cif.			
Assekrem no. 1, Atakor, Algeria	AJ877293§	V	E1 (3)
Ilennanene no. 8, Issekrâm, Algeria	AJ877294‡, AJ877295‡	V	E1 (3)
Jabbârene, Tassili n'Adjer, Algeria	AJ877296‡, AJ877297‡	V, I	E1 (3)
Jebel Marra' crater, Sudan Republic*	AM162275§	I	E1 (7)
Subsp. *cuspidata* (Wall. ex G. Don) Cif.			
Almihwit no. 3, Yemen	AJ878722‡, AJ878731‡	I	C2 (28)
Amalundu, Zimbabwe	AJ877301‡, AJ877302‡	IV	A (37)
Cape Town, South Africa	AJ877300‡	IV	A (39)
China no. 1 (China)	AJ878730§	I	C1 (25)
Gebel Elba, Wadi Darween, Egypt*	AJ878723‡, AJ878724‡	I	E1 (8)
Kerman no. 2, Iran	AJ878726‡	VI	C1 (24)
Kirstenbosch no. 4, South Africa	AJ877298‡, AJ877299‡	IV	A (35)
Guangzhou, ornamental tree no. 1, China	AJ878728‡, AJ878729‡	I	C1 (34)
Mt Elgon no. 1, Kenya	AJ877306‡, AJ877307‡	I	A (36)
Motula, Ethiopia*	AJ877308§	I	C1 (33)
Quetta, Pakistan*	AJ878727§	VI	C1 (24)
Reunion Island no. 1, France	AJ877303‡	IV	A (38)
Teetaem, Oman*	AJ878725§	I	C1 (26)
Timau no. 2, Mt Kenya, Kenya	AJ877304‡, AJ877305‡	I, IV	A (39)

Corresponding ITS-1 and plastid DNA lineages are given. EMBL accession number of each ITS-1 haplotype is indicated.
*Herbarium sample.
†Plastid haplotype number in brackets.
‡ITS-1 sequence after cloning.
§Direct sequencing.

every 100 generations. Topologies prior to Ln-likelihood stabilization ('burn-in'; 5000 generations) were discarded and posterior clade probabilities (pp) were computed from the remaining trees using the 'sumt' command with the 'contype = halfcompat' option. These analyses were replicated three times to ensure topological convergence and homogeneity of posterior probabilities (Huelsenbeck *et al.*, 2002). These phylogenetic analyses were performed on the complete data and then only with sequences generated both by Hess *et al.* (2000) and in the present study (see below).

RESULTS

Distribution of plastid DNA polymorphism

A total of 40 plastid haplotypes was found, as defined based on molecular variation in ccmp5, ccmp7, *trn*T-L, *trn*S-G and *mat*K fragments (Table 1), which greatly extends the number of haplotypes (15) previously identified (Besnard *et al.*, 2002). Ten of the 25 new haplotypes were determined from 28 herbarium samples collected in remote geographical areas. Two main lineages restricted to specific areas can be recognized from the reduced median network (Fig. 1): one corresponding to populations of subsp. *cuspidata* from southern Asia, eastern and southern Africa (lineage cp-I); and a second group, including samples of the other five subspecies from northern Africa, Macaronesia and the Mediterranean plus south-east Egyptian samples of subsp. *cuspidata* (lineage cp-II; Fig. 2). Within lineage cp-I, three haplotype sublineages are recognized: A (from the Rift Valley to southern Africa); C1 (from eastern Africa to south-eastern Asia); and C2 (western Arabia and eastern Africa). Ornamental trees from China display one haplotype (34) belonging to sublineage C1. Within lineage cp-II, four haplotype sublineages are identified: M (Agadir mountains, Macaronesia); E1 (Mediterranean basin, the Sahara); and E2 and E3 (the western Mediterranean) (for lineage coding see Besnard *et al.*, 2002). Africa displays the greatest diversity in haplotype number (33 of 40). Interestingly, we detected 16 haplotypes and three sublineages (E1, E2, M) of subspp. *maroccana*, *guanchica*, *cerasiformis* and *europaea*, distributed in north-west Africa (from the eastern Maghreb to Macaronesia) and 11 haplotypes and four sublineages (A, C1, C2, E1) of subspp. *cuspidata* and *laperrinei*, distributed in north-east Africa. In contrast to multiple haplotypes and lineages in a single area, we observed that wild populations separated by thousands of kilometres share the same haplotypes (e.g. haplotypes 35 and 39 in South Africa and Kenya; 24 in Oman and India; 1 in Syria and Spain; 11 in Libya and Spain).

Characterization of ITS-1 sequences and phylogeny

We isolated 52 ITS-1 sequences from 38 individuals (Table 2). Cloning was required in 17 cases because multiple copies were detected by direct sequencing (see above). All

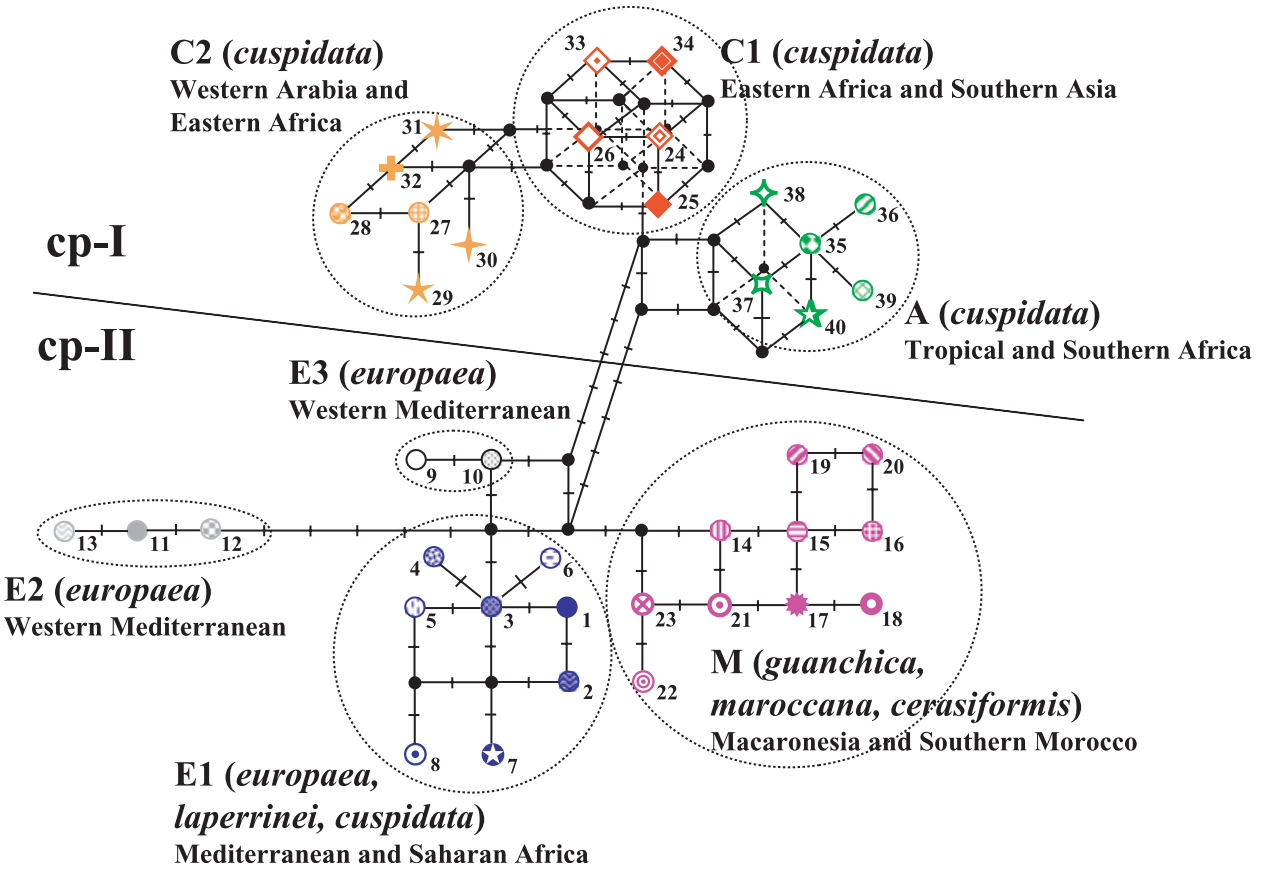

Figure 1 Reduced median network (Bandelt *et al.*, 1999) representing phylogenetic relationships of plastid DNA haplotypes in the *Olea europaea* complex. Each haplotype is represented by a symbol and its corresponding number. Relationships between haplotypes are based on 21 variable characters (Table 1), of which 20 are parsimony-informative. The two lineages cp-I and cp-II are indicated. Main haplotype groupings (sublineages) are named following Besnard *et al.* (2002) (see Results). Their geographical distribution is given (see Fig. 2). Missing, intermediate nodes (small black points) are indicated.

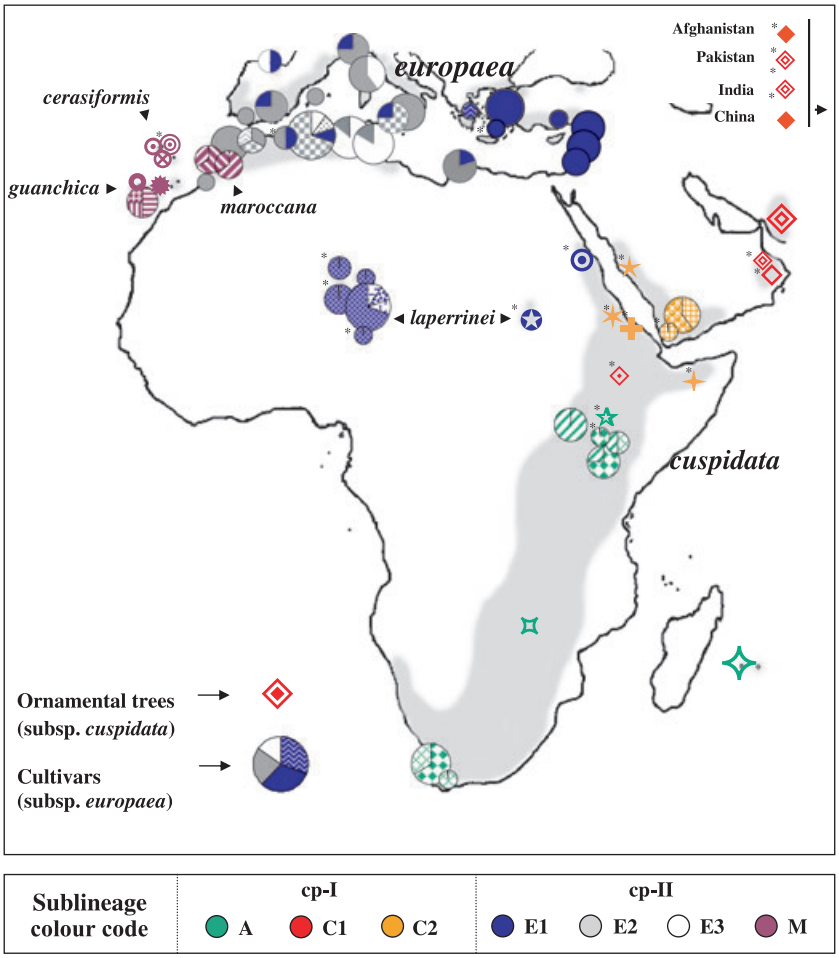

Figure 2 Geographical distribution of the *Olea europaea* complex (shaded area) and plastid DNA haplotypes. Size of pie charts and symbols indicates number of individuals sampled from each location and correspondence with haplotype number in Fig. 1. Individuals characterized using herbarium specimens are indicated by an asterisk. Each plastid sublineage is distinguished by a specific colour.

Table 3 Number of variable and parsimony-informative characters and main features of the maximum parsimony analyses performed on two ITS-1 data sets: total matrix and sequences of group 1

Character	Total ITS-1 matrix (*Olea europaea* + outgroups)	ITS-1 copies of group 1 (*O. europaea*)
Number of variable positions (%)	77.6	34.5
Number of parsimony-informative characters (%)	61.7	23.8
Maximum sequence divergence (GTR)*	6.4	0.25
Informative indels	8	1
Number of most-parsimonious trees	452	824
Tree length	579	138
CI (excluding autapomorphies)	0.514	0.681
RI	0.807	0.911

*GTR distance = general time-reversible distance.

sequences displayed high nucleotide identity (between 86.5% and 100%) and were unambiguously aligned. We failed to obtain PCR amplification from *Olea woodiana*, *O. lancea* and *O. capensis*, species circumscribed in section *Ligustroides* Benth. & Hook. This section includes the closest relatives to the olive complex, which are not sexually compatible with *O. europaea* (Green, 2002). Main features of the MP analyses are shown in Table 3. The MP and BI methods gave similar topologies. These analyses allowed identification of a large monophyletic group (1.0 pp, 88% bs; group 1 in Fig. 3a) formed by the 52 sequences obtained here, plus the five sequences from Hess *et al.* (2000). ITS-1 copies amplified using different primers (ITS-A and ITS-C) clustered together in groups 2 and 3, which are closely related to group 1. Two presumed functional sequences (group 4), amplified with the primers ITS1-For and ITS4-Rev (Brown *et al.*, 1996), displayed a basal-most position in the phylogeny (Fig. 3a).

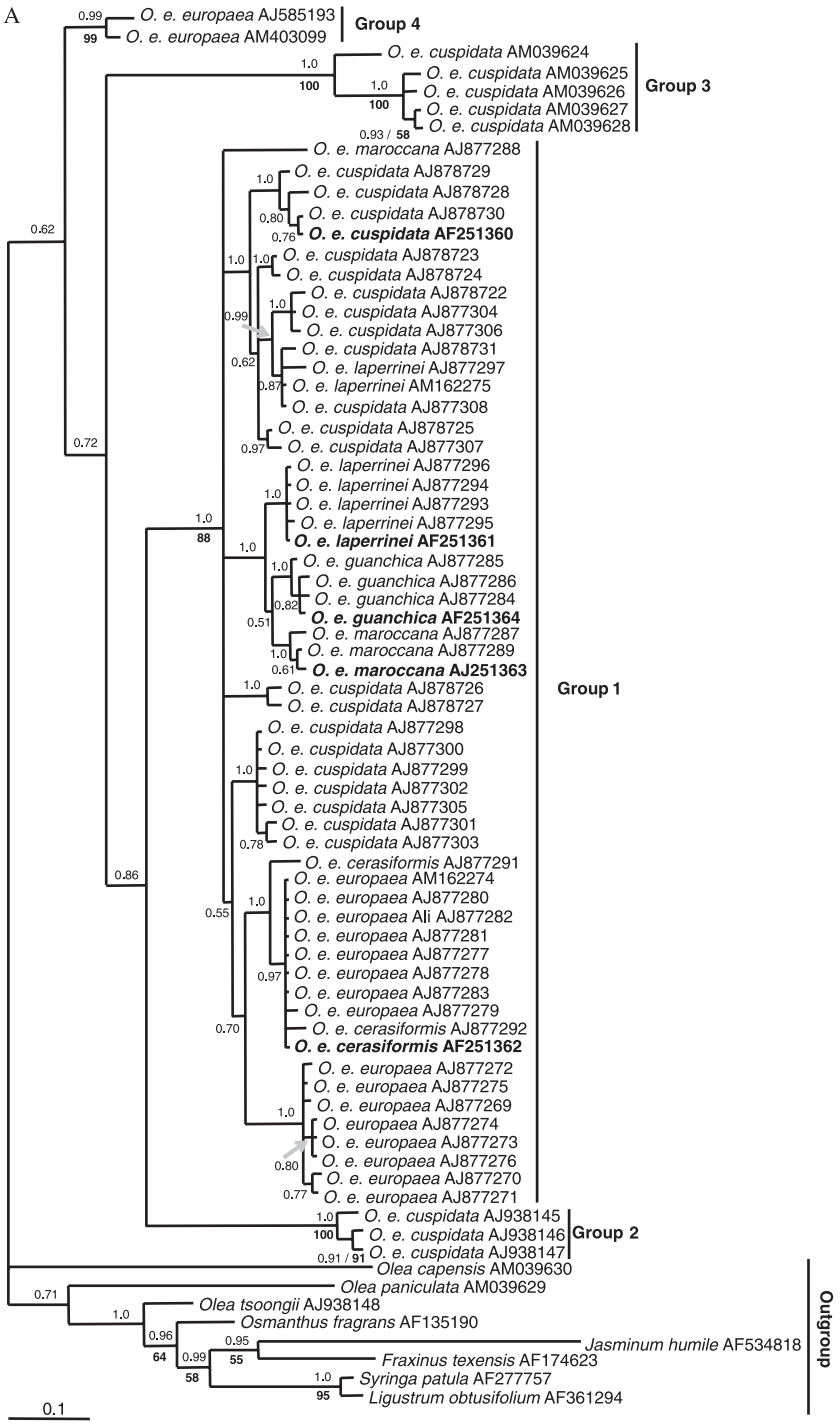

Figure 3 Two 50% majority-rule consensus trees of nrDNA ITS-1 sequences in the *Olea europaea* complex based on a Bayesian inference (BI) approach. The HKY + G model was chosen as the simplest molecular model of evolution (Posada & Crandall, 1998). Support values over 50% are indicated for posterior probability (pp; above) and bootstrap (bs; below) after BI and maximum parsimony analyses, respectively. (A) Phylogeny of ITS-1 sequences in the olive complex. *Olea europaea* accessions are resolved into four groups of closely related sequences (groups 1–4). Sequences forming groups 1–3 correspond to pseudogenes; sequences of group 4 display characteristics expected for functional units (Table 4). Accessions from Hess *et al.* (2000) are in bold. In group 1, bootstrap values are not presented on the corresponding nodes. (B) Phylogenetic relationships of ITS-1 sequences in group 1. Seven main lineages are observed, named I–VII and symbolised. Their geographical distribution is given (see Fig. 4). The tree is rooted to maximize agreement with the tree presented in (a). When two ITS-1 sequences are available for the same individual, accessions are labelled by a specific letter (a–n). Accessions of the four individuals with two divergent ITS-1 sequences are highlighted in bold with a specific colour.

Sequence length, G + C content, minimum-energy secondary structure, sequence similarity with a closely related outgroup (*Osmanthus fragrans*) and nucleotide substitutions at the 5′ ITS starting point and in 5.8S were then investigated in the four ITS-1 groups (Table 4). Group 1 includes shorter ITS-1 sequences (213 bp). This size difference was mainly due to a specific 30-bp deletion. The G + C content of sequences displayed great variation between groups. Groups 1–3 display poor G + C content (48–59%), whereas group 4 shows a higher G + C content (70%). The secondary structure of ITS-1 sequences displayed lower stability in groups 1–3 (Δ = 47.8–62.2 kcal mol^{-1}) than in group 4 (Δ = 92.6 and 96.2 kcal mol^{-1}). Group 4 sequences have the lowest sequence divergence with the *Osmanthus* accession, which is considered to represent a functional ITS unit. Additionally, substitution patterns were analysed to look for potential functional constraints at two highly conserved segments, the 5′ ITS-1 starting point and a 38-bp segment of the 5.8S gene. Only

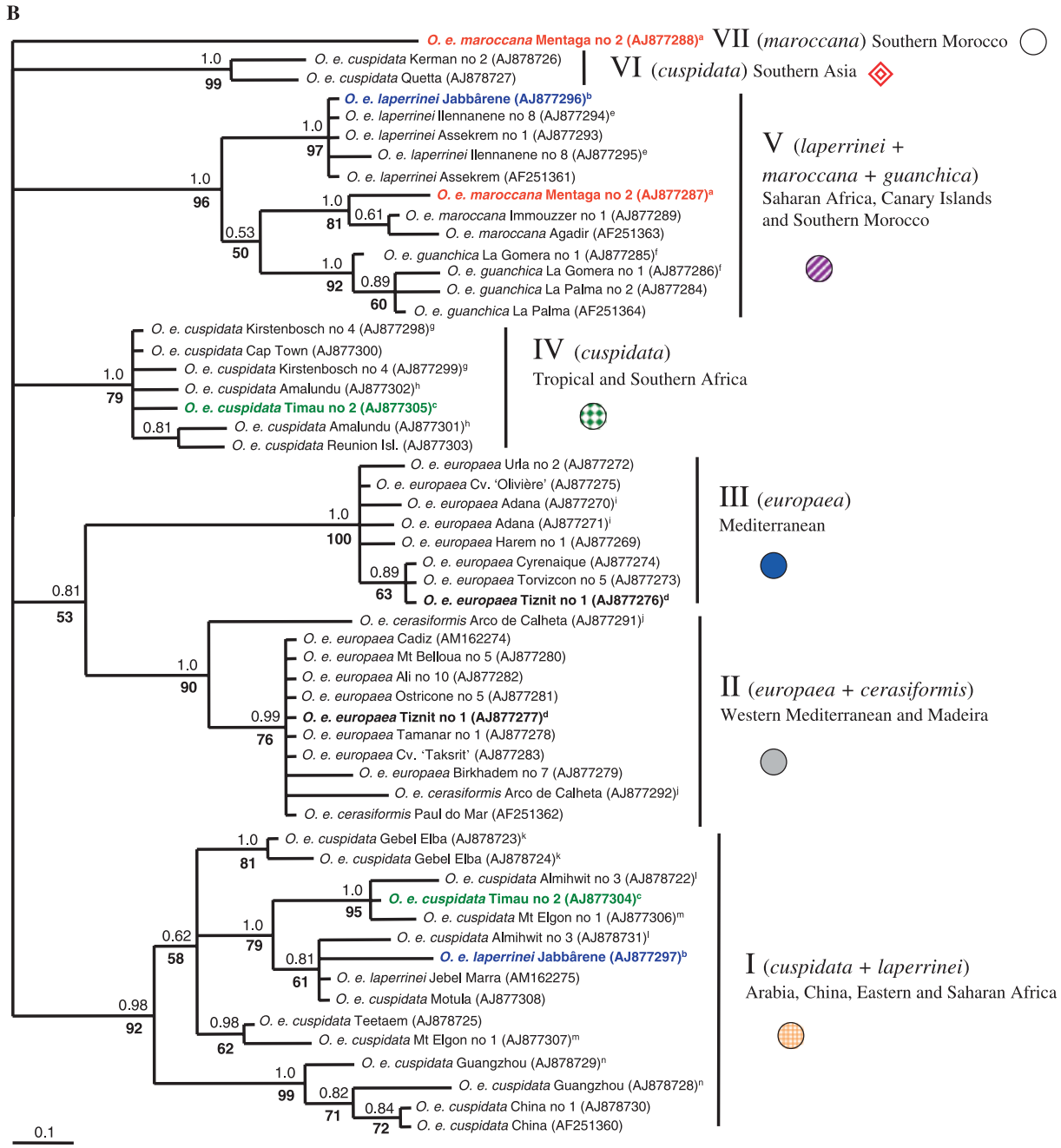

Figure 3b *continued*

sequences belonging to group 4 displayed a low level of substitution in these two segments compared with the outgroup sequence of *Osmanthus*. All these observations support the fact that sequences in groups 1–3 correspond to pseudogenes, while sequences of group 4 display characteristics expected for functional units.

In group 1, seven well supported lineages were recognized in the phylogenetic tree (> 0.98 pp, > 79% bs; Fig. 3b). Each lineage displayed a specific geographical distribution (Fig. 4): lineage I (subspp. *laperrinei*, *cuspidata*) from the Rift Valley to Asia; lineage II (subspp. *europaea*, *cerasiformis*) from Madeira and the western Mediterranean; lineage III (subsp. *europaea*) in all populations analysed from eastern Mediterranean, and in four trees from western Mediterranean (cv. Olivière, Cyrenaique, Torvizcon, Tiznit); lineage IV (subsp. *cuspidata*) from southern Africa to the Rift Valley; lineage V (subspp. *guanchica*, *maroccana*, *laperrinei*) in north-west Africa and the Canary Islands; lineage VI (subsp. *cuspidata*) from two Asian populations (Iran, Pakistan); and lineage VII (subsp. *maroccana*) in a single sample from Mentaga (southern Morocco). Lineages II and III are sister groups (Fig. 3) and therefore may have a more recent divergence than other

Table 4 Main features of the four *Olea europaea* ITS-1 groups (see Fig. 3a)

Character	Group 1	Group 2	Group 3	Group 4
ITS-1 length	213‡	245	244 or 245	245 or 246
G + C content (%)	52.8–59.6	50.6–52.2	48.1–49	68–71
Minimum-energy secondary structure (37°C)*	−47.8 to −61.1‡	−53.6 to −57.5	−52.4 to −62.2	−92.6 to −96.2
GTR sequence divergence with outgroup (*Osmanthus fragrans*, AF135190)	1.09	0.82	1.18	0.38
5′ starting point	TCGA or TCAA‡	TTGA	TCGA	TCGA
Nucleotide-substitution number in 5.8S† (*O. fragrans*, AF135190)	1–7	6–7	5	1

*Minimum-energy secondary structure (Δ) in kcal mol^{-1}.
†Based on 38 bp at 5′ part of 5.8S gene.
‡Characteristics based on the five complete ITS-1 sequences from Hess *et al.* (2000).

lineages. The presence of lineages II, III, V and VII in north-west Africa (including Macatonesia) indicates the highest lineage diversity of *O. europaea* in this area, where four olive subspecies occur. Two distinct ITS-1 sequences were found in 14 individuals. In 10 of these, the two sequences display high similarity (between 95.5% and 99.5%) and were placed in the same ITS-1 lineage (Fig. 3b). Interestingly, four individuals contained two divergent ITS-1 copies of distinct lineages [Tiznit no. 1 (II/III), Jabbârene (I/V), Timau no. 2 (I/IV) and Mentaga no. 2 (V/VII); Table 2; Fig. 3b].

Plastid and nuclear phylogenetic congruence

Cytoplasmic and ribosomal lineages of the 38 individual samples characterized using both plastid and ITS-1 (group 1) polymorphism are summarized in Table 2. Congruent clusters from plastid and nuclear markers (Figs 1 & 3) suggest the following:

1. A remarkable differentiation process occurred in the eastern and western parts of the Mediterranean, with a higher diversity in the western side.

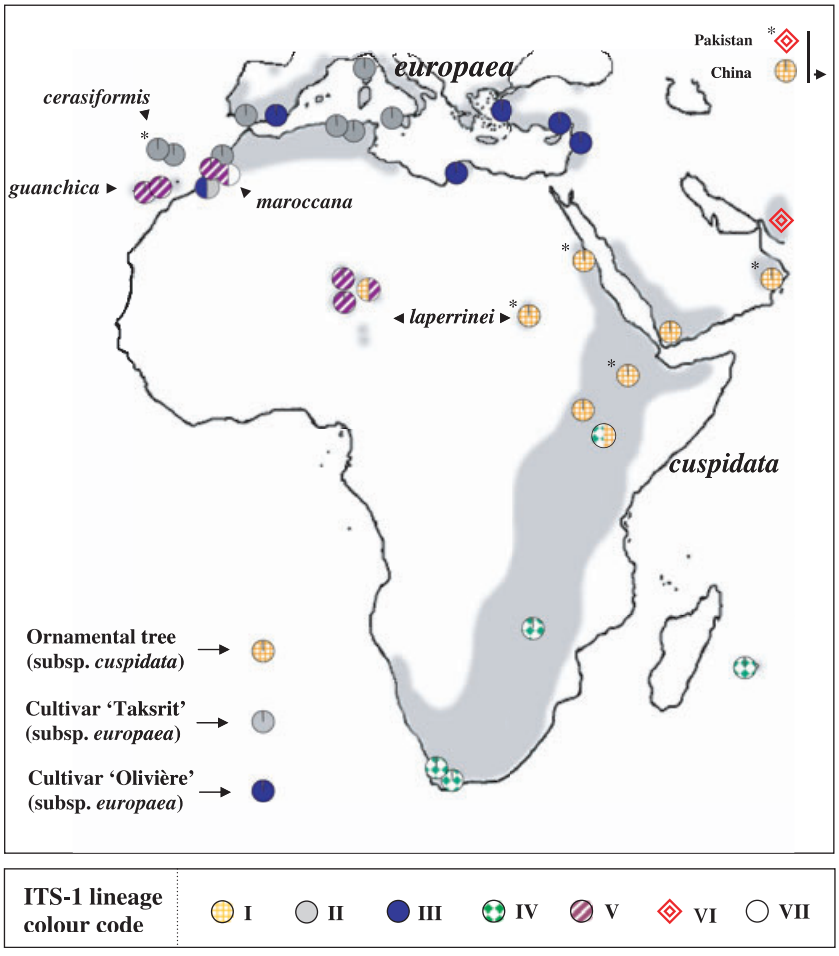

Figure 4 Geographical distribution of the seven ITS-1 lineages of group 1 (see Fig. 3). Individuals characterized using herbarium specimens are indicated by an asterisk. Each ITS-1 sublineage is distinguished by a specific colour.

2. The two sister taxa subspp. *guanchica* and *maroccana* form part of the same plastid and nuclear lineages, which are clearly distinguished from Mediterranean populations.
3. Equatorial and southern African populations of subsp. *cuspidata* are separated from Asian and north-east African populations. With both plastid sublineages and ITS-1 lineages, a transition area is detected between Ethiopia and Kenya (Figs 2 & 4).

Phylogenetic incongruence between the two-genome reconstructions (Figs 1 & 3) was also detected, as follows.
1. In the plastid analysis, subsp. *cerasiformis* is related to *guanchica-maroccana*, while the ITS-1 analysis indicates strong affinities with western Mediterranean populations of subsp. *europaea*.
2. Accessions of subsp. *laperrinei* form part of one (E1) of the three plastid sublineages of Mediterranean populations (*europaea*), but their ITS-1 sequences are closely related to those from both Agadir mountains and the Canary Islands (*guanchica-maroccana*, lineage V) and from north-east Africa (*cuspidata*, lineage I).
3. In subsp. *cuspidata*, the population from Gebel Elba (southern Egypt) displays a plastid lineage primarily distributed in the Mediterranean and the Sahara (sublineage E1), while its ITS-1 sequences are strongly related to those found in both north-east Africa and two populations from the Sahara (lineage I).
4. Two plastid (C1 and C2) and two ITS-1 (I and VI) lineages were found in both South Asia and East Africa (Figs 2 & 4), but the plastid and ribosomal lineage sorting did not lead to a clear identification of two distinct geographical groups.

DISCUSSION

Major lineages of *O. europaea*

The haplotype network based on plastid DNA polymorphism enabled the identification of two main haplotype lineages and seven sublineages (Fig. 1). The clear-cut geographical distribution of lineages cp-I and cp-II supports an early differentiation between populations from southern Africa to China (subsp. *cuspidata*) and the Mediterranean, Macaronesia and the Sahara (all other subspecies) (Vargas & Kadereit, 2001). In this paper, a new western Mediterranean sublineage (E3) and two new closely related sublineages from southern Asia and eastern Africa (C1 and C2) are recognized, in addition to the five described previously (Besnard *et al.*, 2002). Populations of sublineage C1 are present across a large area (from Ethiopia to China; Fig. 2) and may have diverged from an ancient common ancestor that colonized this geographical area. We hypothesize that sublineage C2, formed by populations from African and Arabian coasts of the Red Sea, may reflect a secondary divergence from sublineage C1 (Fig. 1). Haplotype relationships in the sublineage C2 reveals that natural barriers (such as the Red Sea) did not preclude seed exchange between the two continents, in contrast to previous predictions (Besnard *et al.*, 2002).

General distribution of plastid lineages is not fully congruent with present-day taxonomy (Green, 2002). Most of the African individuals of subsp. *cuspidata* display haplotypes from sublineages A, C1 and C2 (lineage cp-I), except for one population from southern Egypt, which forms part of sublineage E1 (lineage cp-II) together with eastern Mediterranean (subsp. *europaea*) and Saharan (subsp. *laperrinei*) populations. In addition, populations of subsp. *europaea* form part of three sublineages (E1, E2, E3; lineage cp-II), while sublineage M (lineage cp-II) indicates a close relationship of three subspecies (subspp. *maroccana*, *guanchica*, *cerasiformis*) of north-west Africa and Macaronesia (Fig. 1). Therefore there are more plastid lineages (7) than recognized taxa, whereas two lineages (E1, M) contain the six subspecies.

Occurrence of divergent ITS-1 paralogues in the olive complex was inferred by phylogenetic analyses (Fig. 3a). Molecular characteristics of available nrDNA sequences (Table 4) indicate that most correspond to non-functional ribosomal copies. Divergent paralogues appear to be responsible for previous difficulties in the analysis of nrDNA in *O. europaea*, as already found in maize (Buckler *et al.*, 1997), *Quercus* (Mayol & Rosselló, 2001), *Ilex* (Manen, 2004) and Rubiaceae (Razafimandimbison *et al.*, 2004). All ITS-1 phylogenetic reconstructions support the monophyly of group 1 (Fig. 3a). In addition, ITS-1 sequences of group 1 are characterized by a specific 30-bp indel. Considering the molecular characteristics, we can assume that these sequences (obtained with a particular primer pair) come from the same nrDNA array. Additionally, molecular analysis of minimum-energy secondary structure, simple models of molecular evolution and comparative features of ITS-1 of each sequence lead us to suggest that group 1 underwent an independent evolutionary process holding phylogenetic signal (Tables 3 & 4). The clustering of taxa within group 1 was basically similar to that previously revealed by multiple fingerprinting markers (Angiolillo *et al.*, 1999; Hess *et al.*, 2000; Besnard *et al.*, 2001b; Rubio de Casas *et al.*, 2006). All these sources of evidence support the utility of the ITS region of the same pseudogene lineage in molecular phylogenetics of the *O. europaea* complex, as shown successfully in other angiosperms (Rauscher *et al.*, 2004; Razafimandimbison *et al.*, 2004). Interestingly, this pseudogenic nrDNA type isolated in all *O. europaea* subspecies was not detected in *Olea* sect. *Ligustroides*, a taxon closely related to the *O. europaea* complex (Baldoni *et al.*, 2002; Green, 2002). The pseudogene identified thus may have been the result of a relatively recent process of evolution post-dating the divergence of *O. europaea* from sect. *Ligustroides*. However, the alternative hypothesis of a lack of sites for ITS-1 primer binding in sect. *Ligustroides* cannot be excluded.

In group 1, each ITS-1 lineage has a limited geographical distribution (Fig. 4), in many cases similar to that of a particular plastid DNA lineage. Both ITS-1 and plastid DNA data indicate that most populations of *O. europaea* are differentiated phylogenetically in five geographical areas: (1)

equatorial and southern Africa; (2) eastern Africa and southern Asia; (3) the eastern Mediterranean; (4) the western Mediterranean; and (5) Macaronesia and north-west Africa. However, unexpected incongruences between plastid DNA and ITS-1 analyses, associated with ITS-1 intra-individual polymorphism (unrelated copies of group 1), led us to propose a more dynamic biogeographical pattern (see below).

Molecular evidence for recurrent hybridization

A high number of lineages, both in plastid and ribosomal DNA analyses, were found primarily in each of two geographical areas: the western Mediterranean and eastern Africa. The occurrence of several glacial refugia in Macaronesia, Morocco, southern Iberia and the eastern Maghreb, coupled with the introduction of multiple cultivars across the Mediterranean bringing about gene flow with local populations, may account for recurrent contacts of olive genomes resulting in a higher genetic diversity observed in the western Mediterranean basin (Besnard *et al.*, 2002; Lumaret *et al.*, 2004). In eastern Africa, the detection of phylogenetically unrelated plastid DNA haplotypes (cp-I, sub-lineages C1, C2 and A; and cp-II, sublineage E1; Fig. 2) and of two major ITS-1 lineages (I and IV; Fig. 4) also supports the hypothesis of secondary contacts between differentiated populations (Besnard *et al.*, 2002). A discordant pattern between plastid DNA haplotypes, ITS-1 sequences and taxonomy has traditionally been considered under the hypothesis of hybridization (Rieseberg *et al.*, 1996; Wendel & Doyle, 1998). The fact that single individuals contain more than one ITS-1 copy of group 1 in four cases shows that concerted evolution failed to homogenize nrDNA from divergent lineages of *O. europaea*. Co-occurrence of two divergent ITS-1 copies in an inter-subspecies hybrid was also reported in an invasive Australian population of *O. europaea* (G.B. and co-workers, unpublished data), which provides further evidence that mixture of divergent copies in a single individual is the result of hybridization between genetically distant genomes.

The body of knowledge accumulated for olive trees in the past decade, including the viability of F_1 crosses between subspecies (Besnard *et al.*, 2001b; G.B., unpublished data), leads us to infer four major reticulation processes:

1. introgression of eastern olives into western Mediterranean populations (reported by Besnard *et al.*, 2001b, 2002);

2. gene flow between Macaronesian and western Mediterranean populations as a result of natural introductions (as suggested by Hess *et al.*, 2000) or human-induced introgression with cultivated olives (as suggested by Green, 2002);

3. secondary contacts in a potential, large hybrid zone consisting of the Saharan mountains (from the Hoggar to southern Egypt) when climatic conditions were favourable; and

4. gene exchange in the Kenyan and Ethiopian highlands (Rift Valley) through a corridor connecting eastern and southern African areas.

The two latter reticulation processes have not been reported previously, and are highlighted in the present study.

Overall, mid-term differentiation of isolated populations (subspp. *laperrinei*, *maroccana*, *guanchica* and *cerasiformis*), coupled with a more recent pattern of hybridization, may account for limited morphological differentiation and incongruence between taxonomy, nuclear sequences and plastid haplotypes in the olive tree complex. Historical isolation and reticulation may have been caused largely by geographical barriers (such as seas or deserts), corridors and climatic changes resulting in waxing-and-waning processes during the Pliocene and Pleistocene.

The Saharan desert: barrier or hybridization arena?

Climatic changes, such as a drier climate in eastern Africa and the Sahara marking the end of the Pliocene (Quézel, 1978; Bonnefile, 1983), are hypothesized to have contributed to the isolation and genetic differentiation of olive populations in the Mediterranean, southern Africa and Arabia. Several species display a geographical distribution similar to the olive tree. *Erica arborea*, and some Andropogoneae species, such as *Andropogon distachyos* and *Hyparrhenia hirta*, are present in the Mediterranean basin, Macaronesia, Saharan mountains, tropical Africa and Arabia, which indicates a shared history of plant distribution between Africa and the Mediterranean (Quézel, 1978). Molecular data have also demonstrated that plant migrations from the Mediterranean to the Saharan mountains occurred in *Senecio* (Coleman *et al.*, 2003), *Atriplex* (Ortíz-Dorba *et al.*, 2005) and *Erica* (McGuire & Kron, 2005). Palaeobotanical data led several authors to consider *O. europaea* as a floristic element of palaeo-tropical origin that colonized the Mediterranean basin from tropical Africa during the Pliocene (Quézel, 1978; Suc, 1984). Our results shed further light on the role of the Sahara in the evolution and distribution of olive trees. Since its formation (*c.* 7 Ma), the Saharan desert may have been an effective barrier, limiting significant reproductive contacts between the two major plastid lineages, which are separated in all molecular analyses (Vargas & Kadereit, 2001; Baldoni *et al.*, 2002; Besnard *et al.*, 2002; Figs 1 & 2). Based on maternally inherited markers (plastid DNA), we hypothesize that significant isolation has been maintained primarily during episodes of the Plio-Pleistocene desertification of the Sahara. We also obtained strong support for more recent gene flow between Mediterranean and subtropical populations, as individuals from Saharan mountains share common nrDNA and plastid DNA lineages with Mediterranean, north-western Africa and eastern African populations. Olive distribution was wider in northern Africa during favourable periods as late as in the Last Glacial Maximum (Maley, 2000). A succession of humid transitional events in the Saharan mountains may have facilitated recurrent contacts between *O. europaea* populations from the Mediterranean basin and tropical Africa. Long-distance gene flow via pollen dissemination may have occurred. Additionally, we also hypothesize that seed dissemination actively mediated by birds, such as pigeons, warblers, blackcaps or blackbirds (Wickens, 1976; Herrera, 1995; Hess *et al.*, 2000; Alcántara & Rey, 2003),

has favoured long-distance dispersal in relatively recent times, in spite of the Saharan desert and Mediterranean sea barriers, as shown by the sharing of certain terminal haplotypes between South Africa and Kenya (haplotype 39), northern Iberia and Tunisia (haplotype 9), or southern France and Libya (haplotype 11) (Figs 1 & 2). Palaeobotanical and palaeoclimatic evidence agrees with a deeper genetic and geographical isolation of the Saharan populations of subsp. *laperrinei* on high mountains as a result of aridification, particularly after the last deglaciation (Quézel, 1978; Gasse *et al.*, 1990). The exceptional survival of *laperrinei* individuals and a low population turnover (Baali-Cherif & Besnard, 2005) may also be responsible for the maintenance of an nrDNA hybridization signal during the past thousands of years.

The phylogenetic patterns found in *O. europaea* can be complemented with phylogeographical patterns of the olive fly (*Bactrocera oleae*), which reveal a Quaternary split of Mediterranean and African populations (Nardi *et al.*, 2005). The Saharan desert again appears to have played an important role in recurrent isolation of organisms, not only in recent times, but also in aridification episodes of the Pliocene and Pleistocene (Quézel, 1978). Molecular-clock estimates of a diphyletic tree of elephant shrews suggest that Miocene climatic events conferred on the Saharan desert the role of vicariant barrier for an even earlier animal isolation (Douady *et al.*, 2003). Miocene and Pleistocene periods also coincide with the historical separation of European and African *Erica* species (McGuire & Kron, 2005). Therefore a long history of climatic shifts in northern Africa triggered a waning-and-waxing distribution of plants in the Sahara territory (Maley, 2000). In addition to wild olives, the occurrence of secondary hybrid zones in eastern Africa was also detected for the honeybee (Franck *et al.*, 2001) and cattle (Hanotte *et al.*, 2002), two other species of agronomic importance. Palaeobotanical data predict a long corridor for plants encompassing the mild eastern Sahara and the mountainous eastern Africa since the uplift of the Ethiopian Highlands during the early Miocene (McGuire & Kron, 2005). An increasing number of phylogenetic and phylogeographical studies on different plant groups will serve to test hypotheses of recurrent isolation by the Saharan desert alternating with hybridization events in favourable climatic periods.

Origin of the cultivated olive

Our plastid DNA and nrDNA results do not support the hypothesis of a recent olive establishment in the Mediterranean from tropical regions (Oliver, 1868; Newberry, 1937; Green, 2002). Uniqueness of lineages and great molecular diversity of plastid and nuclear markers lead us to hypothesize a long-standing occurrence of the olive in the Mediterranean, a scenario also supported by palaeobotanical evidence (Terral *et al.*, 2004b). Common ancestry of *O. europaea* in Africa may have spawned new lineages in Macaronesia, Asia and the Mediterranean long before the Quaternary (Palamarev, 1989). Evolution of the Mediterranean climate in the past 3 Myr may have caused differentiation of subsp. *europaea* populations in the thermophilous vegetation of the Mediterranean basin (Suc, 1984; Palamarev, 1989). Significantly later, Mediterranean olive cultivation began in several locations in relation to Holocene bioclimatic changes (Zohary & Hopf, 2000; Terral *et al.*, 2004a). Archaeological evidence reveals that the cultivated olive was initially domesticated from eastern Mediterranean stocks (Zohary & Hopf, 2000). Today, high levels of genetic and morphological diversity are the result of subsequent hybrid origin of multiple cultivars across the western Mediterranean (Besnard *et al.*, 2001a; Contento *et al.*, 2002). Recurrent hybridization events are probably also related to significant morphological changes observed in archaeological material (Terral *et al.*, 2004a). The hypothesis of a continuous enrichment of olive trees with multiple wild Mediterranean genomes in the course of the domestication is compatible with the molecular similarity of wild and cultivated olives from the same country using multiple markers (Belaj *et al.*, 2001; Besnard *et al.*, 2001a; Vargas & Kadereit, 2001).

ACKNOWLEDGEMENTS

This work was financially supported by the Spanish Ministerio de Ciencia y Tecnología (REN2001-1502-C03-03), European Union grant (HOTSPOTS) no. MEST-CT-2005-020561 and the European Community's programmes BIODIBERIA A-82 and SYNTHESYS ES-TAF-244 at the Royal Botanic Garden of Madrid (CSIC). We thank G. Evanno, N. Salamin, C. Parisod, N. Galland and two anonymous reviewers for their helpful comments on this manuscript, and E. Cano for laboratory assistance. Herbarium samples were provided by P. S. Green (Kew Gardens, Richmond, K), P. A. Schäfer (Montpellier University II; MPU) and R. Spichiger (Geneva Botanical Gardens; G).

REFERENCES

Ainouche, M.L. & Bayer, R.J. (1997) On the origins of the tetraploid *Bromus* species: insights from ITS sequences of nrDNA. *Genome*, **40,** 730–743.

Alcántara, J.M. & Rey, P.J. (2003) Conflicting selection pressures on seed size: evolutionary ecology of fruit size in a bird-dispersed tree, *Olea europaea*. *Journal of Evolutionary Biology*, **16,** 1168–1176.

Álvarez, I. & Wendel, J.F. (2003) Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution*, **29,** 417–434.

Angiolillo, A., Mencuccini, M. & Baldoni, L. (1999) Olive genetic diversity assessed using amplified polymorphic fragment length polymorphisms. *Theoretical and Applied Genetics*, **98,** 411–421.

Armelagos, G.J. & Harper, K.N. (2005) Genomics at the origins of agriculture, part one. *Evolutionary Anthropology*, **14,** 68–77.

Baali-Cherif, D. & Besnard, G. (2005) High genetic diversity and clonal growth in relict populations of *Olea europaea*

subsp. *laperrinei* (Oleaceae) from Hoggar, Algeria. *Annals of Botany (London)*, **96,** 823–830.

Bailey, C.D., Carr, T.G., Harris, S.A. & Hughes, C.E. (2003) Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. *Molecular Phylogenetics and Evolution*, **29,** 435–455.

Baldoni, L., Guerrero, C., Sossey-Alaoui, K., Abbott, A.G., Angiolillo, A. & Lumaret, R. (2002) Phylogenetic relationships among *Olea* species based on nucleotide variation at a non-coding chloroplast DNA region. *Plant Biology*, **4,** 346–351.

Baldwin, B.G., Sanderson, M.J., Porter, J.M., Wojciechowski, M.F., Campbell, C.S. & Donoghue, M.J. (1995) The ITS region of nuclear ribosomal DNA. A valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden*, **82,** 247–277.

Bandelt, H.J., Forster, P. & Röhl, A. (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16,** 37–48.

Belaj, A., Trujillo, I., de la Rosa, R., Rallo, L. & Gimenez, M.J. (2001) Polymorphism and discrimination capacity of randomly amplified polymorphic markers in an olive germplasm bank. *Journal of the American Society for Horticultural Science*, **126,** 64–71.

Besnard, G., Khadari, B., Villemur, P. & Bervillé, A. (2000) Cytoplasmic male sterility in the olive (*Olea europaea* L.). *Theoretical and Applied Genetics*, **100,** 1018–1024.

Besnard, G., Baradat, P., Breton, C., Khadari, B. & Bervillé, A (2001a) Olive domestication from structure of wild and cultivated populations using nuclear RAPDs and mitochondrial RFLPs. *Genetics, Selection, Evolution*, **33**(Suppl. 1), S251–S268.

Besnard, G., Baradat, P., Chevalier, D., Tagmount, A. & Bervillé, A. (2001b) Genetic differentiation in the olive complex (*Olea europaea*) revealed by RAPDs and RFLPs in the rRNA genes. *Genetic Resources and Crop Evolution*, **48,** 165–182.

Besnard, G., Khadari, B., Baradat, P. & Bervillé, A. (2002) *Olea europaea* (Oleaceae) phylogeography based on chloroplast DNA polymorphism. *Theoretical and Applied Genetics*, **104,** 1353–1361.

Besnard, G., Rubio de Casas, R. & Vargas, P. (2003) A set of primers for length and nucleotide-substitution polymorphism in chloroplastic DNA of *Olea europaea* L. (Oleaceae). *Molecular Ecology Notes*, **3,** 651–653.

Bonnefile, R. (1983) Evidence for a cooler and drier climate in the Ethiopian uplands towards 2.5 Myr ago. *Nature*, **303,** 487–491.

Breton, C., Tersac, M. & Bervillé, A. (2006) Genetic diversity and gene flow between the wild olive (oleaster, *Olea europaea* L.) and the olive: several Plio-Pleistocene refuge zones in the Mediterranean basin suggested by simple sequence repeats analysis. *Journal of Biogeography*, **33,** 1916–1928.

Brown, A.E., Sreenivasaprasad, S. & Timmer, L.W. (1996) Molecular characterisation of slow-growing orange and key lime anthracnose strains of *Colletotrichum* from citrus as *C. acutatum*. *Phytopathology*, **86,** 523–527.

Buckler, E.S., IV, Ippolito, A. & Holtsford, T.P. (1997) The evolution of ribosomal DNA: divergent paralogues and phylogenetic implications. *Genetics*, **145,** 821–832.

Chat, J., Jáuregui, B., Petit, R.J. & Nadot, S. (2004) Reticulate evolution in kiwifruit (*Actinidia*, Actinidiaceae) identified by comparing their maternal and paternal phylogenies. *American Journal of Botany*, **91,** 736–747.

Chevalier, A. (1948) L'origine de l'olivier cultivé et ses variations. *Revue Internationale de Botanique Appliquée et d'Agriculture Tropicale*, **28,** 1–25.

Coleman, M., Liston, A., Kadereit, J.W. & Abbott, R.J. (2003) Repeat intercontinental dispersal and Pleistocene speciation in disjunct Mediterranean and desert *Senecio* (Asteraceae). *American Journal of Botany*, **90,** 1446–1454.

Comes, H.P. & Abbott, R.J. (2001) Molecular phylogeography, reticulation, and lineage sorting in Mediterranean *Senecio* Sect. *Senecio* (Asteraceae). *Evolution*, **55,** 1943–1962.

Contento, A., Ceccarelli, M., Gelati, M.T., Maggini, F., Baldoni, L. & Cionini, P.G. (2002) Diversity of *Olea* genotypes and the origin of cultivated olives. *Theoretical and Applied Genetics*, **104,** 1229–1238.

Cronn, R. & Wendel, J.F. (2004) Cryptic trysts, genomic mergers, and plant speciation. *New Phytologist*, **161,** 133–142.

Cronn, R., Small, R.L., Haselkorn, T. & Wendel, J.F. (2003) Cryptic repeated genomic recombination during speciation in *Gossypium gossypioides*. *Evolution*, **55,** 2475–2489.

Douady, C.J., Catzeflis, F., Raman, J., Springer, M.S. & Stanhope, M.J. (2003) The Sahara as a vicariant agent, and the role of Miocene climatic events, in the diversification of the mammalian order Macroscelidea (elephant shrews). *Proceedings of the National Academy of Sciences USA*, **100,** 8325–8330.

Doyle, J.J., Doyle, J.L., Rauscher, J.T. & Brown, A.H.D. (2003) Diploid and polyploid reticulate evolution throughout the history of the perennial soybeans (*Glycine* subgenus *Glycine*). *New Phytologist*, **161,** 121–132.

Elant, H. (1976) Olive. *Evolution of crop plants* (ed. by N.W. Simmonds), pp. 219–221. Longman, London.

Elbaum, R., Melamed-Bessudo, C., Boaretto, E., Galili, E., Lev-Yadun, S., Levy, A.A. & Weiner, S. (2006) Ancient olive DNA in pits: preservation, amplification and sequence analysis. *Journal of Archaeological Science*, **33,** 77–88.

Franck, P., Garnery, L., Loiseau, A., Oldroyd, B.P., Hepburn, H.R., Solignac, M. & Cornuet, J.M. (2001) Genetic diversity of the honeybee in Africa: microsatellite and mitochondrial data. *Heredity*, **86,** 420–430.

Gasse, F., Téhet, R., Durand, A., Gibert, E. & Fontes, J.C. (1990) The arid–humid transition in the Sahara and the Sahel during the last deglaciation. *Nature*, **346,** 141–146.

Green, P.S. (2002) A revision of *Olea* L. (Oleaceae). A revision of *Olea* L. (*Oleaceae*). *Kew Bulletin*, **57,** 91–140.

Green, P.S. & Wickens, G.E. (1989) The *Olea europaea* complex. *The Davis & Hedge Festschrift* (ed. by K. Tan), pp. 287–299. Edinburgh University Press, Edinburgh, UK.

Hanotte, O., Bradley, D.G., Ochieng, J.W., Verjee, Y., Hill, E.W. & Rege, J.E. (2002) African pastoralism: genetic imprints of origins and migrations. *Science*, **296,** 336–339.

Herrera, C.M. (1995) Plant–vertebrate seed dispersal systems in the Mediterranean: ecological, evolutionary, and historical determinants. *Annual Review of Ecology and Systematics*, **26,** 705–727.

Hess, J., Kadereit, J.W. & Vargas, P. (2000) The colonization history of *Olea europaea* L. in Macaronesia based on internal transcribed spacer 1 (ITS-1) sequences, randomly amplified polymorphic DNAs (RAPD), and inter simple sequence repeats (ISSR). *Molecular Ecology*, **9,** 857–868.

Huelsenbeck, J.P., Larget, B., Miller, R.E. & Ronquist, F. (2002) Potential applications and pitfalls of Bayesian inference of phylogeny. *Systematics Biology*, **51,** 673–688.

Kovarik, A., Pires, J.C., Leitch, A.R., Lim, K.Y., Sherwood, A.M., Matyasek, R., Rocca, J., Soltis, D.E. & Soltis, P.S. (2005) Rapid concerted evolution of nuclear ribosomal DNA in two *Tragopogon* allopolyploids of recent and recurrent origin. *Genetics*, **169,** 931–944.

Liao, D. (2003) *Concerted evolution. Encyclopedia of the human genome*. Macmillan/Nature Publishing Group, London. http://www.med.ufl.edu/anatomy/faculty/liao/A0132-Nature-EHG.pdf.

Lumaret, R., Ouazzani, N., Michaud, H., Vivier, G., Deguilloux, M.F. & Di Giusto, F. (2004) Allozyme variation of oleaster populations (wild olive tree) (*Olea europaea* L.) in the Mediterranean Basin. *Heredity*, **92,** 343–351.

Maley, J. (2000) Last glacial maximum lacustrine and fluviatile formations in the Tibesti and other Saharan mountains, and large-scale climatic teleconnections linked to the activity of the subtropical jet stream. *Global and Planetary Change*, **26,** 121–136.

Manen, J.F. (2004) Are both sympatric species *Ilex perado* and *Ilex canariensis* secretly hybridizing? Indication from nuclear markers collected in Tenerife. *BMC Evolutionary Biology*, **4,** 46.

Mayol, M. & Rosselló, J.A. (2001) Why nuclear ribosomal DNA spacers (ITS) tell different stories in *Quercus*. *Molecular Phylogenetics and Evolution*, **19,** 167–176.

McGuire, A.F. & Kron, K.A. (2005) Phylogenetic relationships of European and African Ericas. *International Journal of Plant Sciences*, **166,** 311–318.

McKinnon, G.E., Jordan, G.J., Vaillancourt, R.E., Steane, D.A. & Potts, B.M. (2004) Glacial refugia and reticulate evolution: the case of the Tasmanian eucalypts. *Philosophical Transactions of the Royal Society of London Series B. Biological Sciences*, **359,** 275–284.

Nardi, F., Carapelli, A., Dallai, R., Koderick, G.K. & Frati, F. (2005) Population structure and colonization history of the olive fly, *Bractocera oleae* (Diptera, Tephritidae). *Molecular Ecology*, **14,** 2729–2738.

Newberry, P.E. (1937) On some African species of the genus *Olea* and the original home of the cultivated olive tree. *Proceedings of the Linnean Society of London*, **150,** 3–16.

Oliver, D. (1868) *Flora of tropical Africa*. L. Reeve & Co. Ltd, Lloyds Bank.

Ortíz-Dorba, J., Martínez-Mora, C., Correal, E., Simón, B. & Cenis, J.L. (2005) Genetic structure of *Atriplex halimus* populations in the Mediterranean Basin. *Annals of Botany (London)*, **95,** 827–834.

Owen, C.C., Bita, E.C., Banilas, G., Hajjar, S.E., Sellionakis, V., Aksay, U., Hepaksoy, S., Chamoun, R., Talhook, S.N., Metzidakis, I., Hatzopoulos, P. & Kalaitzis, P. (2005) AFLP reveals structural details of genetic diversity within cultivated olive germplasm from the Eastern Mediterranean. *Theoretical and Applied Genetics*, **110,** 1169–1176.

Palamarev, E. (1989) Paleobotanical evidences of the Tertiary history and origin of the Mediterranean sclerophyll dendro-flora. *Plant Systematics and Evolution*, **162,** 93–107.

Petit, R.J., Duminil, J., Fineschi, S., Hampe, A., Salvini, D. & Vendramin, G.G. (2005) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology*, **14,** 689–702.

Posada, D. & Crandall, K.A. (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics*, **14,** 817–818.

Posada, D. & Crandall, K.A. (2001) Intraspecific phylogenetics: trees grafting into networks. *Trends in Ecology and Evolution*, **16,** 37–45.

Provan, J., Soranzo, N., Wilson, N.J., Goldstein, D.B. & Powell, W. (1999) A low mutation rate for chloroplast microsatellites. *Genetics*, **153,** 943–947.

Quézel, P. (1978) Analysis of the flora of Mediterranean and Saharan Africa. *Annals of the Missouri Botanical Garden*, **65,** 479–534.

Rauscher, J.T., Doyle, J.J. & Brown, A.H.D. (2004) Multiple origins and nrDNA internal transcribed spacer homologue evolution in the *Glycine tomentella* (Leguminosae) allopolyploid complex. *Genetics*, **166,** 987–998.

Razafimandimbison, S.G., Kellogg, E.A. & Bremer, B. (2004) Recent origin and phylogenetic utility of divergent ITS putative pseudogenes: a case study from Naucleeae (Rubiaceae). *Systematic Biology*, **53,** 177–192.

Rieseberg, L.H., Whitton, J. & Linder, C.R. (1996) Molecular marker incongruence in plant hybrid zones and phylogenetic trees. *Acta Botanica Neerlandica*, **45,** 243–262.

Ritz, C.M., Schmuths, H. & Wissemann, V. (2005) Evolution by reticulation: European dogroses originated by multiple hybridization across the genus *Rosa*. *Journal of Heredity*, **96,** 4–14.

Ronquist, F. & Huelsenbeck, J.P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19,** 1572–1574.

Rubio de Casas, R., Besnard, G., Schoenswetter, P., Balaguer, L. & Vargas, P. (2006) Extensive gene flow blurs phylogeographic but not phylogenetic signal in *Olea europaea* L. *Theoretical and Applied Genetics*, **113,** 575–583.

Schaal, B.A., Hayworth, D.A., Olsen, K.M., Rauscher, J.T. & Smith, W.A. (1998) Phylogeographic studies in plants: problems and prospects. *Molecular Ecology*, **7,** 465–474.

Suc, J.P. (1984) Origin and evolution of the Mediterranean vegetation and climate in Europe. *Nature*, **307,** 429–432.

Swofford, D.L. (2001) PAUP. Phylogenetic analysis using parsimony. Version 4. Sinauer Associates, Sunderland, MA, USA.

Terral, J.F., Alonso, N., Capdevila, R.B.I., Chatti, N., Fabre, L., Fiorentino, G., Marinval, P., Jorda, G.P., Pradat, B., Rovira, N. & Alibert, P. (2004a) Historical biogeography of olive domestication (*Olea europaea* L.) as revealed by geometrical morphometry applied to biological and archaeological material. *Journal of Biogeography*, **31,** 63–77.

Terral, J.F., Badal, E., Heinz, C., Roiron, P., Thiebault, S. & Figueiral, I. (2004b) A hydraulic conductivity model points to post-Neogene survival of the Mediterranean olive. *Ecology*, **85,** 3158–3165.

Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22,** 4673–4680.

Turrill, W.B. (1951) Wild and cultivated olives. *Kew Bulletin*, **3,** 437–442.

Vargas, P. & Kadereit, J.W. (2001) Molecular fingerprinting evidence (ISSR, inter-simple sequence repeats) for a wild status of *Olea europaea* L. (Oleaceae) in the Eurosiberian North of the Iberian Peninsula. *Flora*, **196,** 142–152.

Vargas, P., McAllister, H.A., Morton, C., Jury, S.L. & Wilkinson, M.J. (1999) Polyploid speciation in *Hedera* (Araliaceae): phylogenetic and biogeographic insights based on chromosome counts and ITS sequences. *Plant Systematics and Evolution*, **219,** 165–179.

Weising, K. & Gardner, R.C. (1999) A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. *Genome*, **42,** 9–19.

Wendel, J.F. & Doyle, J.J. (1998) Phylogenetic incongruence: window into genome history and molecular evolution. *Molecular systematics of plants II: DNA sequencing* (ed. by D.E. Soltis, P.S. Soltis and J.J. Doyle), pp. 265–296. Kluwer Academic, Boston, MA, USA.

Wickens, G.E. (1976) Speculations on long distance dispersal and the flora of Jebel Marra, Sudan Republic. *Kew Bulletin*, **31,** 105–150.

Zeder, M.A., Emshwiller, E., Smith, B.D. & Bradley, D.G. (2006) Documenting domestication: the intersection of genetics and archaeology. *Trends in Genetics*, **22,** 139–155.

Zohary, D. & Hopf, M. (2000) *Domestication of plants in the Old World*, 3rd edn. Clarendon Press, Oxford, UK.

Zuker, M. (1989) On finding all suboptimal foldings of an RNA molecule. *Science*, **244,** 48–52.

SUPPLEMENTARY MATERIAL

The following supplementary material is available for this article:

Table S1 List and locality of the 28 *Olea europaea* herbarium samples analysed in this study.

Table S2 Geographic origin of the *Olea europaea* individuals sampled in the field and characterized for plastid DNA polymorphism.

This material is available as part of the online article from:http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-2699.2006.01653.x (This link will take you to the article abstract).

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

BIOSKETCHES

Guillaume Besnard (UNIL) currently works on the phylogeography of the wild olive tree. He also studies the post-glacial recolonization of alpine plants and the evolution of adaptive traits (such as heavy metal tolerance and C_4 photosynthesis).

Rafael Rubio de Casas (UCM) works on the phylogeography and evolutionary ecology of Mediterranean plants. In particular, his research focuses on the evolutionary patterns of sclerophyllous taxa.

Pablo Vargas (Royal Botanic Garden of Madrid) leads projects on the evolution and molecular systematics of primarily Mediterranean plant groups. He is currently investigating the biogeography, phylogeography, speciation and microevolution of Western Mediterranean and Macaronesian genera.

Editor: Robert Whittaker